

READ

Repository for Electronic Aggregation of Documents

CS 410

Fall 2012

“Black” Team



The READ Team

Project Leader
Marcus Zehr

Risk Management
Jacob Philmon

Database Specialist
Jim Calderon

Software Designer
Troy Connor

Domain Expert
Dr. Michele Weigle

Database Specialist
Andrew Sprague

Software Designer
Andrew Moss

Scraper Algorithm
Andrew Schaefer

Scraper Algorithm
Satish Lakkoju

Scraper Algorithm
Chinmay Lokesh

Outline

- Problem statement
- Problem
 - Case Study
 - Case Study issues
- Problem solution
- Main benefits
- Customer base
- What it will and won't do
- Competing solutions
- Risks
 - Technical Risks
 - Consumer Risks
- Conclusion
- References

Problem

Organizations lack an efficient method for storing their member's **publications, research grants**, and other documents, as well as displaying them to the general public or the rest of the organization

Societal Effects

- Public Universities get 20% of their income for federal grants and contracts ¹
- Universities have some difficulty sharing ideas freely between them
- There is a lack of publicity and recognition of work that otherwise would be valued

¹freeby50.com

ODU Case Study

The ODU Computer Science Department lacks an efficient and up-to-date system for housing, displaying, searching and sorting department publications and grants. Only one person was in charge of running the system



Overview of Case Study

- Adding grants or publications is a manual process
- Grants and publications awarded and/or published by ODU personnel are not stored in a centralized location
- Individuals who are interested in these grants or publications do not know where to look for this information
- Stored documents only include tech reports, grants, and publications
- Anyone may view items submitted to the system, but only a few individuals can submit them
- Faculty are not searchable and are hard to recognize for their work

Current System Issues

How is it updated

- Past: Citations were manually submitted by email and entered on the site manually
 - Slow, inefficient, and manual
- Present: System is currently outdated and the most recent publication is from 2008

Searching Issues and System Access

- Anyone can view the current system
- Currently unorganized
- Only sorted by year
- No way to filter the static page
- Not accessible via CS website

Old Dominion UNIVERSITY | A to Z Index | Directories | Search ODU | GO

College of Sciences

COMPUTER SCIENCE

Recent Publications

Graduate

- ▶ PhD Program Information
- ▶ MS Program Information
- ▶ Course Information

Undergraduate

- ▶ Program Information
- ▶ Course Information
- ▶ Advising Information

Student Resources

- ▶ Admissions
- ▶ Financial Aid
- ▶ Fall 2012 Schedule
- ▶ Fall 2012 Books
- ▶ Research Opportunities
- ▶ Employment Opportunities
- ▶ Calendars
- ▶ Appointed Students
- ▶ ODU Association of Computing Machinery

Directory

- ▶ Faculty
- ▶ Staff
- ▶ Current Student and Alumni

Systems Group

- ▶ Main Page
- ▶ Computing Lab Hours
- ▶ High Performance Computing Services
- ▶ Virtual Computing Lab
- ▶ Networking Services
- ▶ Obtaining Microsoft Software through ELM's
- ▶ CS IT Infrastructure 2012 ([pdf](#))
- ▶ CS Services Writeup ([docx](#))

J. Crouch, Y. Shen, J. Austin, M. Dinniman, An Educational Interactive Numerical Model of the Chesapeake Bay," *Computers & Geosciences*, Vol. 34, No. 3, pp. 247-258, **March 2008.**

K. Maly, P. Flynn, L. Zhou, S. Zeil, M. Zubair, "Automated Template-Based Metadata Extraction Architecture," *Asian Digital Libraries: Looking Back 10 Years and Forging New Frontiers, Lecture Notes in Computer Science*. Vol. 4822, No. 2007, pp. 326-336, **January 2008.**

F. McCown, J. A. Smith, M. L. Nelson, J. Bollen, Reconstructing Websites for the Lazy Webmaster, *Dynamics of Search Engines: An Introduction*, pp. 81-112, Icfai University Press, Hyderabad, **2008.**

V. Alturi, S. Chun, R. Mukkamala, P. Mazzoleni, "A Decentralized Execution Model for Inter-Organizational Workflows," *Journal of Distributed Parallel Databases*, Vol. 22, No. 1, pp.55-83, **August 2007.**

K. Bhoopalam, K. Maly, R. Mukkamala, and M. Zubair, "Framework for Information Sharing Across Multiple Government Agencies under Dynamic Access Policies," *Second International Conference on Availability, Reliability and Security*, **2007.**

J. Bollen, M. L. Nelson, G. Geisler, R. Araujo, "Usage Derived Recommendations for a Video Digital Library," *Journal of Network and Computer Applications*, Vol. 30, No. 3, pp. 1059-1083, **August 2007.**

C. Boonthum, Dr. Levinstein, I.B. & McNamara, D.S. In A. Kao & S. Poteet (Eds.) "Evaluating Self-Explanations in iSTART: Word Matching, Latent Semantic Analysis, and Topic Models," *Text Mining and Natural Language Processing*, London, Springer Verlag UK, **2007.**

R.R. Brooks, B. Pillai, M.C. Weigle, and M. Pirretti, Optimal Layout of Multicast Groups Using Network Embedded Multicast Security in Ad Hoc Sensor Networks, *International Journal of Distributed Sensor Networks (IJDSN)*, **July 2007, Vol 3, No 3, pp. 273-287.**

J. Crouch, S. Pizer, E. Chaney, Y. Hu, G. Mageras, M. Zaider, "Automated Finite Element Analysis for Deformable Registration of Prostate Images," *IEEE Transactions on Medical Imaging*, Vol. 26, No. 10, pp. 1379-1390, **October 2007.**

J Crouch, A Cherry. "Parametric Eye Models." *Medicine Meets Virtual Reality (MMVR) 2007*.

K. Ibrahim and M.C. Weigle, Accurate Data Aggregation for VANETs (poster), *Proceedings of ACM VANET*, Montreal, Canada, **September 2007, pp. 71-72.**

www.cs.odu.edu/recent_publications.shtml

The READ Solution

Database integration

- MySQL
 - Store documents and links
 - Relatively easy to setup, use and maintain
 - Inexpensive
 - Easy to pick up and learn
- Web-based interface
- Web-based Scraper
 - Aggregate specific publication databases only
- User logins and password protection



Web-based Scraper

- A **scraper** is a software tool that searches websites for publications
- Used to scrape online databases that currently exist with publications already uploaded
- Run at a set interval that is set by the consumer
- Can be turned on or off depending if a separate online database exists for the consumer
- Much of it has already been completed by Andrew Schaefer

Web-based Interface

- Allows searching of the database using keywords
- Users will login to upload and edit links to their publications

The screenshot shows a web-based interface for searching publications. At the top, there is a navigation bar with links for "Publications | Grants | Faculty" and a "Login" button. Below the navigation bar, there is a search form with the following elements:

- Keywords:** A text input field containing the text "TextBox".
- Start Date:** A date input field.
- End Date:** A date input field.
- Order by:** A dropdown menu.
- Full Text Available:** A dropdown menu.
- Items Shown Per Page:** A text input field.
- Filter:** A button to execute the search.

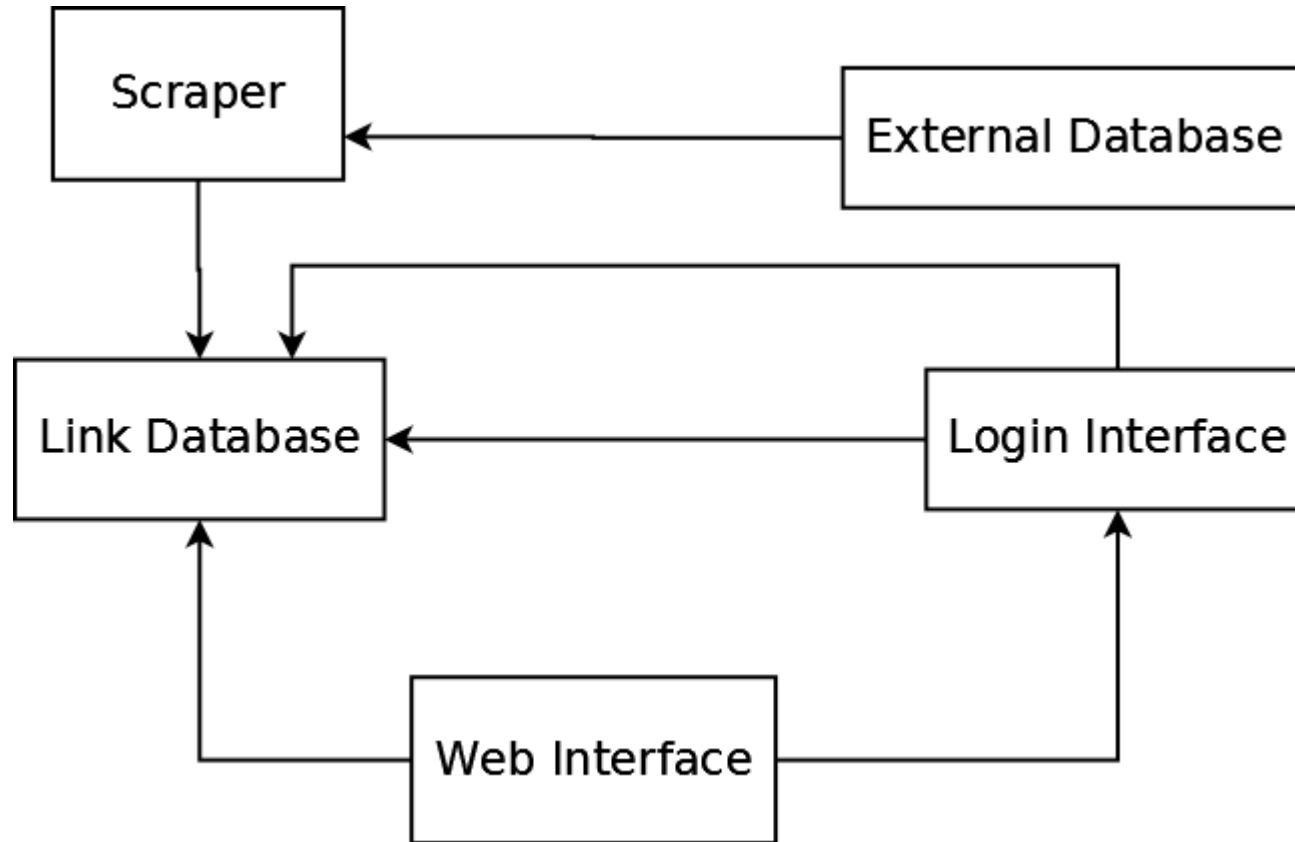
Web-based Interface

- Users may also add items that they may wish to associate with their publications such as video files, thumbnails, and presentations
- Access to documents and the system is set by the consumer for the generic problem
- Users will be able to submit thumbnail images that will be displayed next to their publications when displayed in the search interface

Publication Moderation

- Documents will be moderated by the users which upload them
- An administrator will also have control over documents

Major Component Diagram



Main Benefits

- Allows for an organized and simple way to navigate for publications, grants, and tech reports
- Provide an effective method to publicize the research that has been done at ODU's CS Department for grant distribution organizations
- Provide an effective method to advertise ODU's CS Department to future students
- Preserves information

What It Does

What It Will Do

- It will provide a method to easily access ODU's CS research and grants
- It will search the internet for research already done by faculty

What It Won't Do

- Provide access to copyrighted or protected material
- Gather research material from anyone outside of the university

Potential Customers

The people who would take interest in this project are:

- 673 Undergraduates, 111 Master Students, 11 PhD in ODU CS Dept. 37 Staff
- 24.4 thousand at ODU ²
- 24.6 million students enrolled in Fall 2011
- 4,706 Degree Granting institutions ³

²<http://www.odu.edu/>

³http://nces.ed.gov/programs/digest/d11/tables/dt11_001.asp?referrer=report

Competition Matrix

	READ	Mobius	Eprints	SPL	irplus	eScholarship
Information Offered						
Publications	X	X	X	X	X	X
Grants	X				X	
Site Features						
Scraper	X					
Searchable	X	X	X	X	X	X
Filterable	X	X		X	X	X
Sortable	X	X	X	X	X	X
Thumbnail Images	X					
Author Uploads	X		X	X	X	X
Abstracts	X	X		X	X	
Statistics	X					
Reference Formats						
Bibtex	X	X	X	X		
Google Scholar	X	X				
XML	X	X	X	X		
Endnote			X			X

		Impact				
		1	2	3	4	5
Probability	5					
	4				TR 1	CR 3
	3			TR 2	TR 3	
	2		CR 1			
	1	CR 2				

Technical Risks:

TR1) System Interface may be incompatible with specific computer devices

TR2) The interface may be incompatible with certain browsers as a whole

TR3) SQL injection

Consumer Risks:

CR1) Users with aged, or hardcopy publications may want to have them included in the database

CR2) Certain parties may not be willing to use the interface due to not wanting to use technology or understanding how it works

CR3) Legal aspects: some associated items the user wishes to submit may be illegal to upload onto our system (IEEE documents for example)

Technical Risk 1

1) System Interface may be incompatible with specific computer devices(mobile phone, ipad, computer, etc.)

- Mitigation: Possibly create a mobile page or maybe even an app to interface with the database for the devices that need to access it

Probability: 4 Original Impact: 4 Shift: 0

Technical Risk 2

- 2) The interface may be incompatible with outdated browsers or even certain browsers as a whole
- Mitigation: Make specific versions for incompatible browsers
 - Problems: Larger workload
- Probability:3 Original Impact: 3 Shift: 0

Technical Risk 3

3) Hackers may try to use SQL injection to forcibly take protected data stored in the database

- SQL injection is the use of SQL commands to extract data out of the database
- Mitigation: Hash and encrypt data at risk
Probability: 3 Original Impact: 4 Impact shift: 1

Consumer Risk 1

1) Users with aged, or hardcopy publications may want to have them included in the database

Mitigations:

- Ask Consumers to scan such documents before uploading
 - Problem: More work on the consumers end

Probability:2

Original Impact: 2

Shift: 1

Consumer Risk 2

2) Certain parties may not be willing to use the interface due to not wanting to use technology or understanding how it works

- Mitigation: Provide a manual or user guide for the system

Probability:1

Original Impact: 1

Shift:0

Consumer Risk 3

3)Legal issues: some documents submitted may not be allowed to be submitted on the system due to copyrights or such (IEEE Documents)

- Mitigation: Have a moderation system implemented (human or automated) that will remove such items and warn the user who submitted it not to do so again

Probability: 4 Original Impact: 5 Impact Shift:1

Conclusion

- READ will allow for the creation of a dynamic web page which is user friendly, easy to update, and easy to maintain
- Applicable for many different organizations
- Information will be easy to find and access
- Added functionality
- Questions



Repository for Electronic Aggregation of Documents

References

Goodin, Dan. "Hackers Expose 453,000 Credentials Allegedly Taken from Yahoo Service." *Ars Technica*. N.p., 12 July 2012. Web. 22 Sept. 2012. <<http://arstechnica.com/security/2012/07/yahoo-service-hacked/>>.

Multi-State Information Sharing and Analysis Center. *Why Cyber Security Is Important*. East Greenbush, NY 12061: Multi-State Information Sharing and Analysis Center, 2005. [Http://wyoming.gov/](http://wyoming.gov/). Oct. 2005. Web. 18 Sept. 2012. <http://wyoming.gov/pdf/brochure_security-is-important.pdf>.

"Publications | MOBIUS." *MOBIUS | Mobius group @ UIUC*. . N.p., n.d. Web. 7 Nov. 2012. <<http://mobius.cs.uiuc.edu/pubs>>.

Year. "CRTC - Publications By Year." *CRTC - Center for Real-Time Computing*. N.p., n.d. Web. 7 Nov. 2012. <http://crtc.wm.edu/html_output/publications_by_year.php>.

References

"Technical Reports Â ." *About Technical Reports*. MIT Libraries, n.d. Web. 25 Sept. 2012. <<http://libguides.mit.edu/techreports>>.

"Publication." *Dictionary.com*. Dictionary.com, n.d. Web. 25 Sept. 2012. <<http://dictionary.reference.com/browse/publication?s=t>>.

"Importance of Academic Publishing." *UNITEN*. UNITEN, n.d. Web. <<http://www.uniten.edu.my/newhome/uploaded/admin/research/publications/2011/On%20the%20Importance%20of%20Academic%20Publishing%20rotated.pdf>>.

"Delta Cost Project Data." *The Delta Cost Project on Postsecondary Education Costs, Productivity, and Accountability*. Delta Cost Project, n.d. Web. 26 Sept. 2012. <<http://www.deltacostproject.org/resources/>>.

References

"Where Do Universities Get their Money From? | Free By 50." *Free By 50*. N.p.,
p.,

n.d. Web. 9 Nov. 2012. <<http://www.freeby50.com/2011/11/where-do-universities-get-their-money.html>>.

"ODU - Old Dominion University." *ODU - Old Dominion University*. N.p., n.d.
Web. 13 Nov. 2012. <<http://www.odu.edu/>>.

"eScholarship | University of California." eScholarship. N.p, n.d. Web.
11 Nov. 2012. <<http://www.escholarship.org/>>

"Eprints - Digital Repository Software." Eprints. N.p, n.d Web.
11 Nov. 2012. <<http://www.eprints.org/>>

References

"irplus - Institutional Repository Software." irplus. N.p, n.d. Web. 11 Nov. 2012. <<http://code.google.com/p/irplus/>>

"PubDB - The Publication database." SPL - Surgical Planning Laboratories. N.p,n.d Web. 11 Nov. 2012. <<http://www.spl.harvard.edu/publications/>>